doi:10.1038/nature10163

Supplementary Table 1 | Efficiency of vector construction.

Process	wells	recovered	efficiency (%)
Recombineering*	480	461	96
Intermediate plasmids†	461	381	83
Recombineering efficiency	**480**	**381**	**80**
Final targeting vectors†	381	363	95
Gateway efficiency	**455**	**363**	**95**
Overall efficiency	**480**	**363**	**76**

*growth in wells; †sequence-verified plasmids

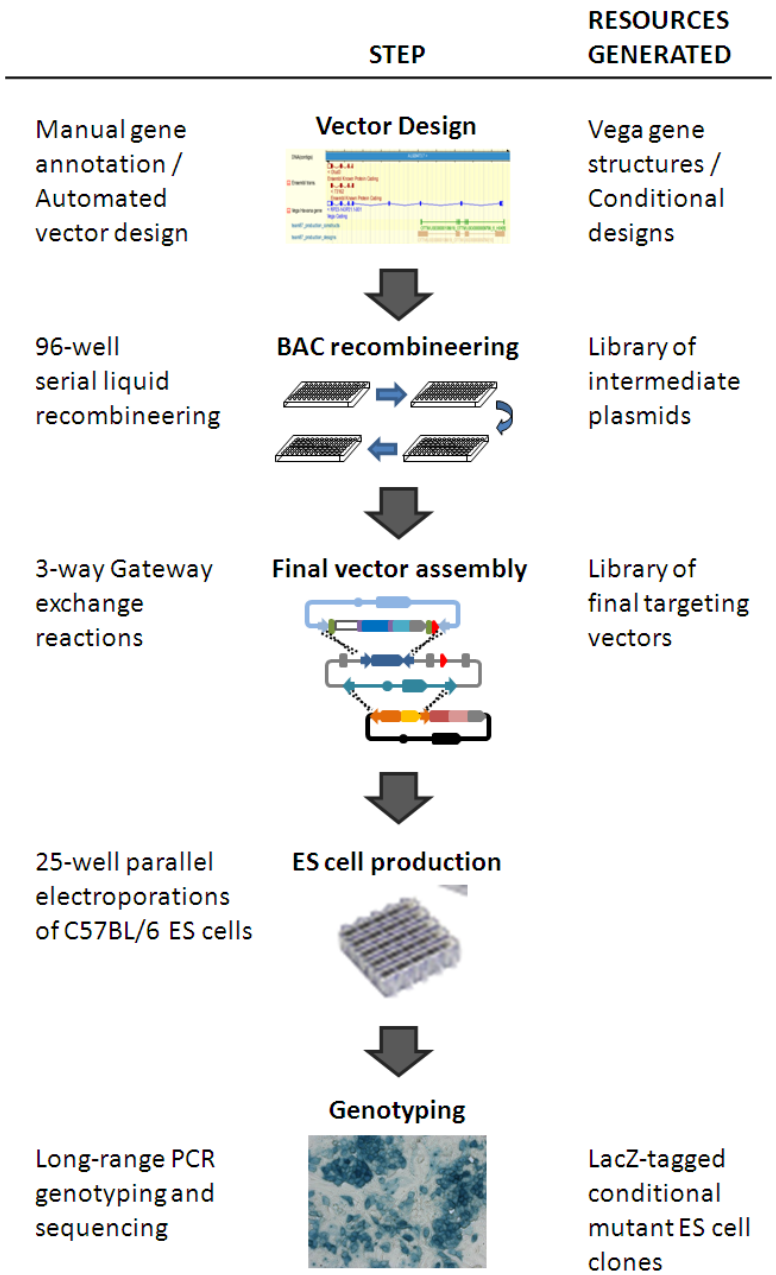
Supplementary Table 2 | Colony number is a predictor of targeting success with promoterless, but not with promoter-driven, targeting vectors.

Vector type	Colony #	# genes	targeting efficiency	success rate
Promoterless	failed	507	n.a.	n.a.
	1 to 50	141	23%	52%
	51 to 100	136	36%	74%
	101 to 200	151	47%	79%
	>200	376	65%	94%
	All	**1285**	**51%**	**48%**
Promoter	failed	140	n.a.	n.a.
	1 to 50	242	35%	79%
	51 to 100	192	33%	85%
	101 to 200	294	30%	85%
	>200	943	37%	89%
	All	**1811**	**35%**	**80%**

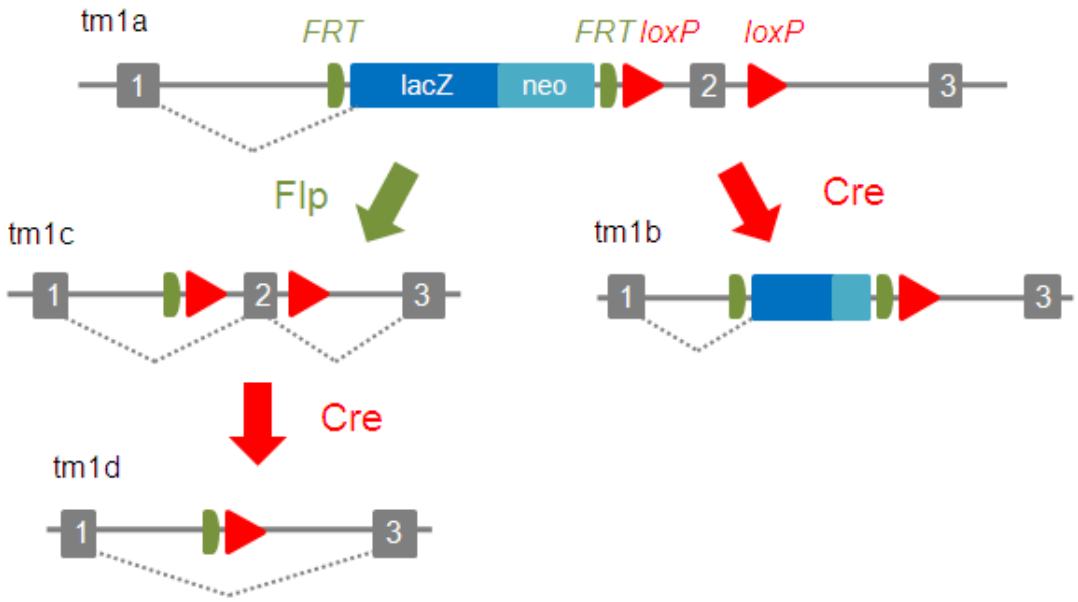
Supplementary Table 3 | Correlation of targeting efficiency with number of gene traps and gene class.

Vector type	# TIGM traps*	# genes	targeting efficiency	success rate	Gene class	# genes	targeting efficiency	success rate
Promoterless	0	582	32%	28%	Secreted	374	36%	28%
	1 to 4	368	52%	57%	Non-secreted	911	54%	57%
	>4	335	65%	74%				
Promoter	0	977	34%	80%	Secreted	670	34%	78%
	1 to 4	544	36%	77%	Non-secreted	1141	35%	80%
	>4	290	34%	83%				

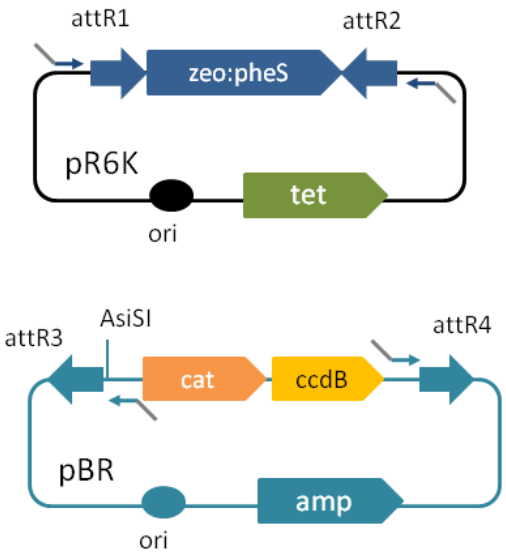
*trapping data from the International Gene Trap Consortium (www.genetrap.org)



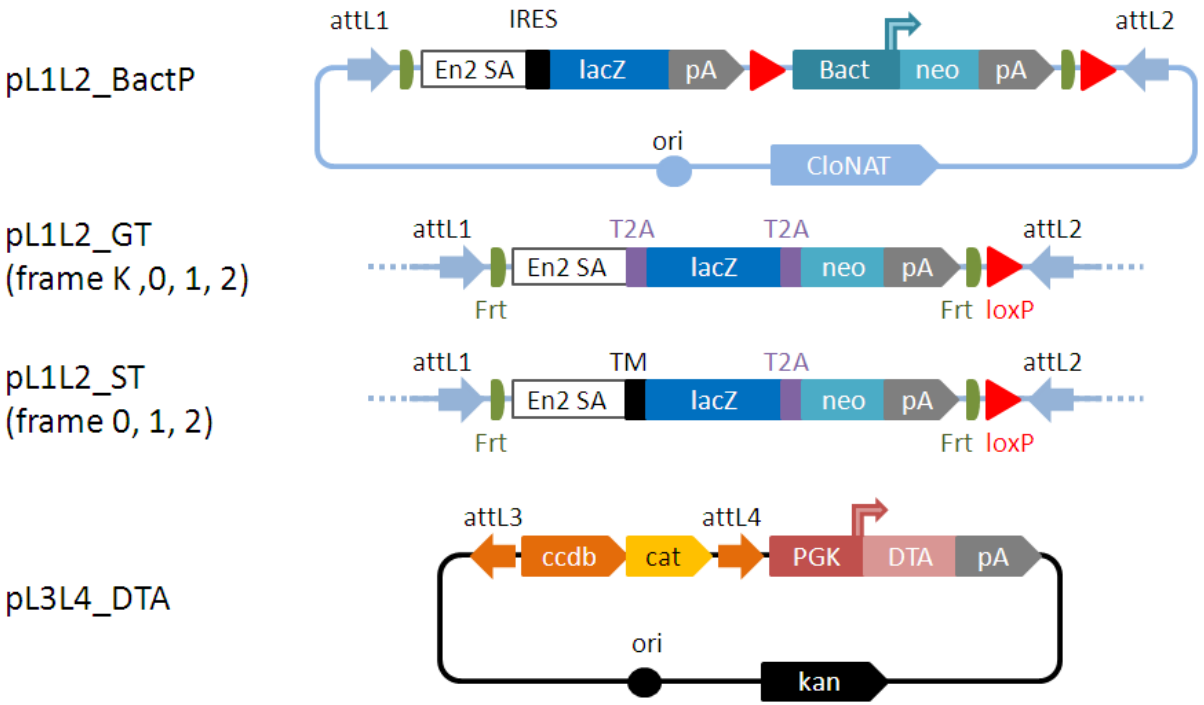
Supplementary Figure 1 | Summary of high-throughput gene targeting pipeline. The process begins with manual Vega annotation of gene structures[1] followed by computational design of conditional alleles. Optimal designs are manually selected and oligos/BACs are ordered in 96-well format for high-throughput serial recombineering. Intermediate Gateway-adapted vectors are then assembled into final targeting constructs in a 3-part *in vitro* reaction. Sequence-verified constructs are used in 25-well parallel electroporations of C57BL/6N ES cells and drug-resistant clones are screened by LR-PCR and sequencing to identify correctly targeted ES cells harboring lacZ conditional alleles.



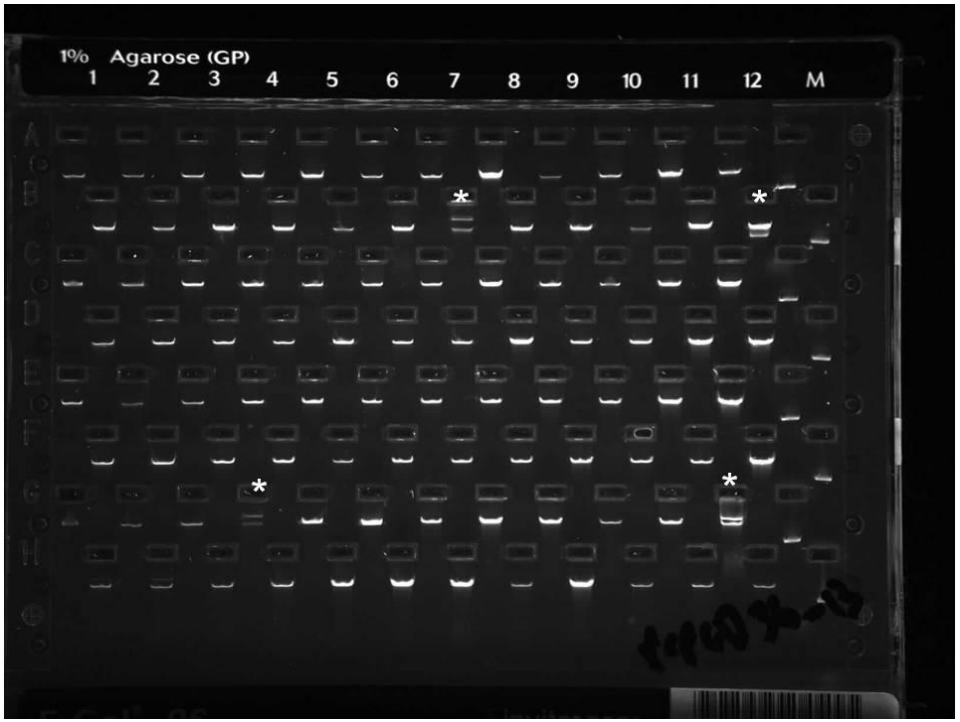
Supplementary Figure 2 | Schematic of the 'knockout-first' allele generated with promoterless vectors. The intital allele (tm1a) contains a *lacZ* promoterless trapping cassette inserted into the intron of a gene, disrupting gene function. Flp recombinase converts the 'knockout-first' allele to a conditional allele (tm1c), restoring gene activity. Cre recombinase deletes the trapping cassette and floxed exon of the tm1a allele to generate a *lacZ*-tagged allele (tm1b) or deletes the floxed exon of the tm1c allele to generate a frameshift mutation (tm1d), triggering nonsense mediated decay (NMD)[2] of the deleted transcript.



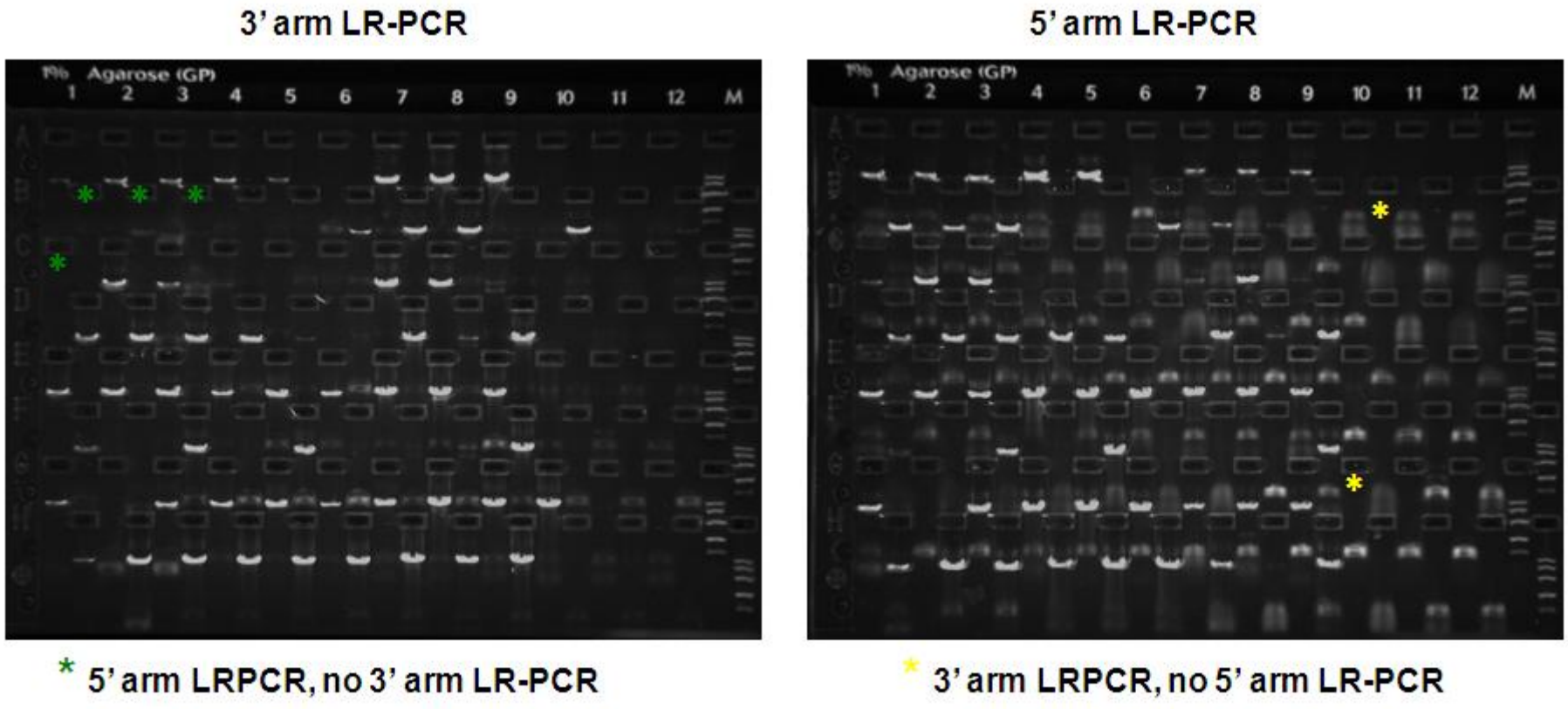
Supplementary Figure 3 | Custom recombineering cassettes for the construction of Gateway-adapted intermediate plasmids. The R1/R2 pheS:zeo recombineering plasmid (top) contains the R6K origin of replication and the bacterial EM7 promoter driving the *zeo* and *pheS* genes. The *pheS* gene[3] serves as a negative selection marker for the Gateway exchange reaction. The attR1 and attR2 sites are each flanked by 20 bp of unique sequence for PCR amplification (arrows). The R3/R4 plasmid backbone for gap repair (bottom) is a modified pBR322 plasmid (*AmpR, TetS*) containing attR3 and attR4 sites flanking a bacterial positive-negative selection cassette (*cat-ccdB*). The R3 and R4 sites are each flanked with 20 bp of unique sequence for PCR amplification (arrows). A very rare AsiSI site next to the R3 site is used to linearize the final targeting construct.



Supplementary Figure 4 | Custom Gateway-adapted targeting elements for the construction of targeted vectors. The promoter-driven targeting cassette (**pL1L2_BactP**) contains IRES:*lacZ*[4] flanked by the En2 splice acceptor sequence and SV40 polyA (pA) signal[5]. A floxed human β-actin promoter[5] (Bact) drives the expression of neo in ES cells. FRT sites are included on either side of the targeting cassette. Promoterless targeting cassettes (**pL1L2_GT**) contain an in-frame fusion of lacZ and neo flanked by the En2 splice acceptor sequence and SV40 polyA (pA) signal[6]. T2A peptide sequences[7] are included to allow expression of lacZ and neo as independent polypeptides. These vectors have been engineered in each of the three possible reading frames (0, 1, 2) and with a Kozak ATG (K)[6]. FRT sites are included on either side of the cassette along with a loxP site at the 3' end of the vector. Secretory trap versions of the promoterless targeting cassettes (**pL1L2_ST**) contain the rat CD4 transmembrane domain (TM) fused to the N-terminus of *lacZ*[6]. The negative selection plasmid backbone (**pL3L4-DTA**) contains L3 and L4 sites flanking the bacterial *ccdB* and *cat* genes followed by a diphtheria toxin A-chain (DTA) expression cassette. DTA is expressed from the mouse phosphoglycerate kinase 1 (PGK) promoter[8].



Supplementary Figure 5 | Gel electrophoresis of final targeting vectors linearized with AsiSI. Most clones migrate as a single high-molecular weight band of the expected size. Occasionally, contaminating DNA fragments of unknown origin are observed (asterix)



Supplementary Figure 6 | E-gels are used to analyze the products of LR-PCR. For most genes, both 5' and 3' arm genotyping yield positive bands of the predicted size (>5 kb) yield and expected sequence. For some clones, positive sequence-verified bands are obtained for only one arm or the other (asterix). Lanes 1-3: Tbc1d10a; Lanes 4-6: Agxt2l2; Lanes 7-9: Prpsap2; Lanes 10-12: C1qbp. DNA markers (M).

References

1. Wilming, L. G. et al. The vertebrate genome annotation (Vega) database. Nucleic Acids Res. 36, D753–D760 (2008).

2. Baker. K.E. & Parker R., Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol. 16, 293-299 (2004).

3. Kast, P. pKSS—a second-generation general purpose cloning vector for efficient positive selection of recombinant clones. Gene 138, 109–114 (1994).

4. Mountford, P. et al. Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression. Proc. Natl Acad. Sci. USA 91, 4303–4307 (1994).

5. Gossler, A., Joyner, A. L., Rossant, J. & Skarnes, W. C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science 244, 463–465 (1989).

6. Skarnes, W. C., Moss, J. E., Hurtley, S. M. & Beddington, R. S. Capturing genes encoding membrane and secreted proteins important for mouse development. Proc. Natl Acad. Sci. USA 92, 6592–6596 (1995).

7. Szymczak, A. L. et al. Correction of multi-gene deficiency in vivo using a single ‘self-cleaving’ 2A peptide-based retroviral vector. Nature Biotechnol. 22, 589–594 (2004).

8. Soriano, P., Montgomery, C., Geske, R. & Bradley, A. Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. Cell 64, 693-702 (1991).